DOCUMENT RESUME

ED 154 399 CS 204 127

TITLE Testing and Evaluation in English and the Language Arts: Abstracts of Doctoral Dissertations Published in "Dissertation Abstracts International," March through December 1977 (Vol. 37 No. 9 through Vol. 38 No. 6).

INSTITUTION ERIC Clearinghouse on Reading and Communication Skills, Urbana, Ill.

PUB DATE 78

NOTE 13p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.

DESCRIPTORS Annotated Bibliographies; Black Students; *Communication Skills; Composition (Literary); *Doctoral Theses; Educational Research; Elementary Secondary Education; *English Instruction; Faculty Evaluation; Higher Education; Language Arts; Language Skills; *Measurement Instruments; Predictive Measurement; Standardized Tests; *Student Evaluation; Test Construction; Test Reliability; Test Validity

ABSTRACT

This collection of abstracts is part of a continuing series providing information on recent doctoral dissertations. The 15 titles include studies dealing with the following areas: the evaluation of student writing; the assessment of the communicative competence of kindergarten children; formulas for predicting high school reading achievement; an English usage examination at a Texas university; a survey of secondary school English teachers' knowledge and attitudes regarding testing and accountability; considerations in designing a composition placement instrument; assessment of the functional communication competence of 17 year old students; the influence of readers on the holistic grading of essays; the construct validity of the Grammatical Closure subtest as a measure of language ability; a test of the critical listening ability of college students; the training of preservice teachers in the use of standardized language arts achievement tests; the effects of the cultural context of language on the performance of black inner-city students; the reliability of ratings on the Language Skills Examination; faculty evaluation procedures for secondary level English teachers; and methods for determining the internal consistency of English compositions. (GW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED154399

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Testing and Evaluation in English and the Language Arts:

Abstracts of Doctoral Dissertations Published in Dissertation Abstracts International, March through December 1977 (Vol.37 No. 9 through Vol. 38 No. 6)

Compiled by the staff of

ERIC/RCS

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Dissertations Abstracts International

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

CS 204 127

The dissertation titles contained here are published
with permission of the University Microfilms International,
publishers of Dissertation Abstracts International
(copyright © 1977 by University Microfilms International),
and may not be reproduced without their prior permission.

This bibliography has been compiled as part of a continuing series designed to make available information on relevant dissertations to users of the ERIC system. Monthly issues of Dissertation Abstracts Interntional are reviewed in order to compile abstracts of dissertations on related topics, which thus become accessible in searches of the ERIC data base. Ordering information for the dissertations themselves is included at the end of the bibliography.

Abstracts of the following dissertations are included in this collection:

Allen, Charles Lee

A STUDY OF THE EFFECT OF SELECTED MECHANICAL ERRORS ON TEACHERS' EVALUATION OF THE NONMECHANICAL ASPECTS OF STUDENTS' WRITING

Black, Janet Knecht

FORMAL AND INFORMAL MEANS OF ASSESSING THE COMMUNICATIVE COMPETENCE OF KINDERGARTEN CHILDREN

D'Angelo, John Louis

PREDICTING READING ACHIEVEMENT IN A SENIOR HIGH SCHOOL FROM INTELLIGENCE, LISTENING, AND INFORMATIVE WRITING

Duncan, William Neil

THE STUDENT USE OF ENGLISH EXAMINATION AT NORTH TEXAS STATE UNIVERSITY: 1944-1976

Infantino, Robert Louis

TESTING AND ACCOUNTABILITY: A SURVEY OF THE KNOWLEDGE AND ATTITUDES OF NEW YORK STATE SECONDARY SCHOOL ENGLISH TEACHERS

Johnson, Nancy Wehrle

THE USES OF GRAMMATICAL AND RHETORICAL NORMS, PEDAGOGICAL STRATEGIES, AND STATISTICAL METHODS IN DESIGNING AND VALIDATING A COMPOSITION PLACEMENT INSTRUMENT

Mead, Nancy Alexander

THE DEVELOPMENT OF AN INSTRUMENT FOR ASSESSING FUNCTIONAL COMMUNICATION COMPETENCE OF SEVENTEEN YEAR OLDS

Newcomb, Joseph Scott

THE INFLUENCE OF READERS ON THE HOLISTIC GRADING OF ESSAYS

Reese, Linda Lee

A STUDY TO DETERMINE THE CONSTRUCT VALIDITY OF THE GRAMMATIC CLOSURE SUBTEST AS A MEASURE OF LANGUAGE ABILITY USING SIGNAL DETECTION THEORY

Richards, Roger Allan

THE DEVELOPMENT AND EVALUATION OF A TEST OF CRITICAL LISTENING FOR USE WITH COLLEGE FRESHMEN AND SOPHMORES

Shank, Kathlene Stapfer

TRAINING PRE-SERVICE TEACHERS IN USE OF STANDARDIZED LANGUAGE ARTS ACHIEVEMENT TEST: AN EXPERIMENTAL PROJECT

Simpkins, Charlesetta

EFFECTS OF THE CULTURAL CONTEXT OF LANGUAGE ON THE COGNITIVE PERFORMANCE OF BLACK STUDENTS

Singleton, Doris Jacobs

THE RELIABILITY OF RATINGS ON THE ESSAY PORTION OF THE LANGUAGE SKILLS EXAMINATION

Sutton, Gary Alan

EVALUATION OF FLORIDA SECONDARY ENGLISH TEACHERS

Wynn, James Howard

DETERMINING THE INTERNAL CONSISTENCY OF ENGLISH COMPOSITIONS USING SELECTED CRITERIA

A STUDY OF THE EFFECT OF SELECTED MECHANICAL ERRORS ON TEACHERS' EVALUATION OF THE NON-MECHANICAL ASPECTS OF STUDENTS' WRITING

ALLEN, Charles Lee, Ed.D.
Temple University, 1976

Problem

The purpose of this study was to investigate the extent to which the presence of mechanical errors in students' writing affects teachers' evaluations of the non-mechanical aspects of the writing. The specific questions investigated in the study concerned: (1) the effect of errors in spelling, punctuation, capitalization, and sentence structure on teachers' evaluation of the quality, organization, and development of an idea and the individual style of the content of students' writing; (2) the relative effect of such errors on junior high school teachers as opposed to senior high school teachers; and (3) the relationship of the varied aspects of students' writing to each other.

Procedures

The subjects of this study were 412 teachers of secondary school English in a large urban school district. The 42 schools involved in the study were randomly assigned to one of four groups so that each group was composed of a cross-section of the district's schools.

Following the identification of a typical sample of uncorrected student writing, the investigator reworte the sample in three different versions. In the first revised version, all mechanical errors were removed. For the second version, the student's original mechanical errors were retained, and several additional ones were supplied by the investigator. For the third version, all mechanical errors were retained, and syntactical patterns of a variety of non-standard English were added by the investigator. In all four versions of the writing sample, the word choice of the student was the same. All four versions of the writing sample were assigned to one of the four groups of teachers.

Each group of teachers was requested to evaluate the paragraph assigned to it using a 1 – 7 rating scale for each of eight categories. Although four of the categories concerned the mechanics of writing and four concerned the content of writing, this division was not brought to the teachers' attention.

Findings

The results of this study have been summarized in terms of the three hypotheses of the investigation. First, there were no significant differences among the mean content ratings for the four versions of the writing sample, between the ratings of junior and senior high school teachers, or among the interaction of the main effects. There were many significant and highly significant correlations, as determined by the Spearman rank correlation coefficient (rho). Among the most consistently significant correlations were those: (1) between the total subscore for mechanics and the total score assigned to the writing sample, and (2) between the total subscore for content and the total score assigned to the writing sample. On all four versions of the writing sample these correlations were all significant at the .001 level.

Conclusions

Subject to the limitations of the study, the following conclusions were drawn: 1. Many teachers are already able to ignore the presence of mechanical errors when they are evaluating only the non-mechanical aspects of students' writing. 2. Teacher-training institutions and school districts need to identify such teachers, determine the criteria they use in evaluation, and design programs to assist other teachers to evaluate writing more effectively. 3. Students who write in a dialect significantly different from Standard English do not suffer unjustly because their writing is judged against the conventions of edited written Standard English. 4. Junior high school teachers are less likely than senior high school teachers to: (1) perceive mechanical errors in writing, or (2) be negatively affected by the presence of mechanical errors. 5. The length of time teachers have been teaching, the number of graduate hours of study in English, and the number of graduate hours of study in education have no relationship to their ability to evaluate either the mechanics or the content of students' writing.

Order No. 76-15,827, 119 pages.

FORMAL AND INFORMAL MEANS OF ASSESSING THE COMMUNICATIVE COMPETENCE OF KINDERGARTEN CHILDREN

BLACK, Janet Knecht, Ph.D.
The Ohio State University, 1977

Adviser: Professor Martha L. King

The assessment of kindergarten children's communicative competence was examined with regard to two formal instruments (the Illinois Test of Psycholinguistic Abilities and Circus 10) and the informal environment of the sociodramatic area.

Communicative competence, as defined in this investigation, has two aspects (1) grammatic competence, and (2) interactional competence. Grammatic competence was measured by the kindergarten child's syntactic maturity, vocabulary diversity, and use of irregular inflections in the three oral language samples.

The second phase of the investigation concerned the assessment of interactional competence. Pinnell's (1975) categorization of the functions of language was applied to the three evaluative approaches to determine the number of functions elicited by each assessment technique. The seven functions include Instrumental, Regulatory, Interactional, Personal, Imaginative, Informative, and Heuristic. In addition, oral language samples collected from the sociodramatic area via two tape recorders, a video tape, and the researcher's notes were analyzed according to the interactional competency checklist. The interactional competency checklist was designed by the researcher and was based upon the thought and research of Cicourel, (1972); Hymes, (1971); Lewis and Cherry, (1975); Mishler, (1976); Ritchie-Key, (1976); and Schegloff and Sacks, (1973). The interactional competency checklist consists of four categories: (1) the ability to adapt to changes in the setting, (2) nonverbal appropriateness, (3) familiarity with normal constraints and conditions of conversation, and (4) sequencing.

Analysis of the data suggests that the formal means of assessment used in this investigation provided limited and inaccurate information about a child's grammatic competence, and provided no information regarding the child's interactional competence. However, the sociodramatic area generally provided more comprehensive information about the child's total communicative competence.

In terms of vocabulary diversity, the sociodramatic area provided kindergarten children with the opportunity to use a significantly greater (.001 level) variety of words than either of the two standardized test situations. The natural environment interview indicated that children could use irregular noun, verb, and pronoun inflections to a significantly greater degree of accuracy (.001 level) than the two standardized tests suggested. There was no significant difference between the three oral language samples in terms of the analysis of syntactic complexity.

The second phase of the investigation concerned the interactional aspect of communicative competence. The nature of the sociodramatic area provided the opportunity for the subjects to use six of the seven functions of language while Circus 10 and ITPA each elicited only one function or use of language. This study also determined that the sociodramatic area of the kindergarten classroom provided an environment in which the researcher was able to document numerous examples of the subjects' interactional competency as categorized by the interactional competency checklist.

This study provides evidence that children have virtually obtained grammatic and interactional competence by the time they enter school. In short, if children are given the opportunity to interact in environments over which they exercise some control, their response tends to be more natural and reflective of their true communicative competence. These findings raise some questions for educational practice concerning the (1) complete dependency upon standardized tests in assessing language performance, (2) the "deficit" approach used in evaluating children's oral language, and (3) the narrow approach of language development programs. It would appear that the opportunity to practice and use language with peers in the natural environment could be the most effective language program.

In conclusion, this study suggests that the assessment of communicative competence, if it is to be reliable, valid and comprehensive, must include interactional as well as grammatic competence. This recommendation is indicative of the trend in language research which takes into consideration a more global kind of competence necessitating semantic analysis in the social setting. Order No. 77-24,598, 221 pages.

PREDICTING READING ACHIEVEMENT IN A SENIOR HIGH SCHOOL FROM INTELLIGENCE, LISTENING, AND INFORMATIVE WRITING

D'ANGELO, John Louis, Ed.D.
University of Pittsburgh, 1977

Adviser: Dr. Harry W. Sartain

The purpose of this study was to develop a regression formula which could be used to predict the reading achievement of high, average, and low ability ninth grade students from knowledge of mental age and selected skills in informational writing and listening. A secondary purpose was to determine the inter-rater reliability of an assessment instrument for informational writing.

The data-collection instruments employed in this study were the Otis-Lennon Mental Ability Test, J and K; the Brown-Carlsen Listening Comprehension Test; the Listening to Remember Specification Test, Level E; and the Informative Writing Scale, Advanced Middle School Level.

A sample total of 245 students was randomly selected from the total ninth grade population of a suburban Pittsburgh school district. The 245 students were categorized as possessing high, average, or low ability on the basis of the mental ability test. Thus, 102 students were selected for the high ability category, 120 were selected for the average ability category, and 23 were selected for the low ability category. The various measuring instrument scores of 77 students in the high category, 90 students in the average category, and 18 students in the low category were utilized in the development of a regression formula for that category. The remaining students in each category served as a cross-validation group to determine the effectiveness of each equation.

After the administration and scoring of the various instruments, correlation coefficients were obtained to determine the relationships between reading achievement and the independent variables. Multiple regression techniques then were employed to obtain predictive formulae and a stepwise technique was used to determine the best predictors. The results of the formulae then were compared with actual reading scores to ascertain the effectiveness of each formula.

All correlation coefficients for the total sample population were significant at the .01 level of confidence with the Otis-Lennon Mental Ability Test having the largest correlation with reading achievement, .67031. However, none of the multiple regression formulae which were found to be most effective used the mental ability variable. All of the formulae which were found to be most effective used a measure of listening comprehension and/or listening memory.

The regression formula developed for the total main sample was found to be more convenient and less time-consuming to use than the other formulae which were developed because students did not have to be grouped according to ability. This formula, which employed student scores on the listening comprehension, listening memory, and mechanics section of the writing scale, was found to have a 79.5 percent of agreement and a 20.5 percent of error when compared with actual reading scores.

The regression formula developed for the low ability group employed only the listening memory score and was found to have an 89 percent of agreement and only an 11 percent error when compared with actual reading scores. This formula was recommended for populations which have been grouped according to ability.

An analysis of variance technique employed to determine if a significant difference existed among three raters using the Informative Writing Assessment Instrument revealed no significant difference at the .01 level of confidence.

Order No. 77-21,216, 131 pages.

THE STUDENT USE OF ENGLISH EXAMINATION AT NORTH TEXAS STATE UNIVERSITY: 1944-1976

DUNCAN, William Neil, Ed.D.
North Texas State University, 1976

This study is concerned with the history of the Student Use of English examination at North Texas State University in Denton, Texas. Conceived in 1944 and implemented in 1946, the examination serves as a measure to insure that marginal English students demonstrate a minimal proficiency in composition evidenced by a three and one-half page essay written on topics related to their major fields. The topics, changed annually, are not disclosed until the time of the supervised examination. A marginal student, one who has made any grade of D in the required four courses in English, must be judged "satisfactory" to graduate from the university.

The examination has undergone four phases, from one in which it was required of all students to the present one in which it is taken by marginal students only. Records show that no less than 22,403 students have written for the examination since its inception.

Analysis of the marked errors on 971 examinations disapproved since 1962 reveals a mean of twenty errors on each composition, with spelling and commas constituting 9.4 errors. No failing paper contained less than eleven cited errors.

An analysis of five hundred questionnaires returned from former students and one hundred questionnaires from current students indicates the following findings: (1) the largest percentage of students taking the examination is from the College of Education, but a larger percentage of failures comes from the College of Business; (2) 48.4 per cent of the former students and thirty-three per cent of the current students spent no time in preparing for the examination; (3) some students hint that others passed the examination dishonestly; (4) fifty-eight

per cent of the current students and 54.4 per cent of the former students feel that the requirement does nothing to enhance the reputation of the university; (5) 45.8 per cent of the former students and fifty-five per cent of the current students favor abandoning the requirement as a prerequisite to graduation; (6) although each of the respondents made at least one D in English, twenty-five per cent state that they do a great deal of writing in their current employment, the largest number of whom are now engaged in teaching; and (7) only twenty per cent of the respondents currently have secretaries.

The major conclusions of the dissertation are that no improvement in writing skills is revealed on the failing examinations for a fourteen-year period, and that 49.7 per cent of the surveyed students feel that the examination is not a learning experience. Though no specific criterion exists for evaluation of examinations, the graders have been consistent in their marking of errors.

The recommendations include the implementation of a new phase in which only students with a cumulative English average of less than C be required to take the examination and a recommendation that the freshman English program place more emphasis upon writing. In addition, a recommendation is made to encourage professors in other disciplines to stress more writing in their subjects.

The essential significance of this dissertation is that it assembles all extant historical documents concerning the examination in one place, it determines a criterion for failing examinations through an analysis of 971 disapproved essays, and it records evaluative responses from six hundred former and current writers of the Student Use of English examination at North Texas State University.

Order No. 77-11,102, 307 pages.

TESTING AND ACCOUNTABILITY: A SURVEY OF THE KNOWLEDGE AND ATTITUDES OF NEW YORK STATE SECONDARY SCHOOL ENGLISH TEACHERS

INFANTINO, Robert Louis, Ed.D.
State University of New York at Buffalo, 1976

Problem

This study was conducted to describe the knowledge and attitudes of secondary school English teachers in New York State toward the uses of externally prepared examinations, including standardized tests and statewide Regents examinations, and toward the effect that accountability programs are having on teachers, school systems, and pupils, especially accountability programs which use standardized testing of pupils as a measure of teacher performance. Twelve research questions were posed for this study.

Procedure

A random sample of 985 public school secondary English teachers was drawn from all school systems in New York State, using the computer resources of the New York State Education Department. These teachers were mailed a survey questionnaire containing thirty-seven items concerned with the teacher's knowledge of current testing practices. An additional twenty-seven questionnaire items were attitudinal, using a five-point Lickert-type scale. On these items teachers were asked to indicate their attitudes toward testing and accountability.

A total of 497 useable questionnaires were returned, a 50.5% rate. Follow-up interviews were conducted with ten randomly selected respondents from the Western New York area.

Findings

Descriptive data were compiled using the program package SPSS. Some of the major findings were: 1. There is a wide variety of all types of standardized tests being administered in New York State. 2. Tests results are accessible to teachers and are used appropriately for guidance and placement purposes; teachers are seldom forced to help students prepare for standardized tests. 3. Younger, less experienced teachers feel restricted in both content and method by externally prepared examinations such as the Regents Comprehensive English Examination. 4. Teachers showed a lack of familiarity with testing terminology and test construction. They did seem to recognize problems inherent in standardized testing such as cultural bias and unreliability of a single test score. 5. Pupil performance on standardized or externally prepared examinations does not affect teacher evaluation, promotion, or tenure. 6. Accountability programs are not widespread in New York State and do not appear to be having much effect on teachers or pupils. 7. Teachers expressed a willingness to be held accountable for those factors over which they have some control such as knowledge of subject matter. They are not willing to be held accountable for bringing about growth in pupil learning based on standardized test scores.

Recommendations

Selected recommendations based on the survey and interview data are these: 1. School districts in New York State should review testing practices to be sure that the standardized tests being administered are adequate in regard to the present curriculum, the population of the district, and the usefulness of the test results to the district. Tests which are not adequate should be discounted. 2. Teachers should not be unnecessarily restricted in content or methodology in English classes because of externally prepared examinations. District-wide and State-wide final English examinations should be eliminated. 3. Teachers should not now be evaluated on the basis of how well or poorly their students perform on standardized tests or externally prepared tests because of inadequate knowledge about the relationship between teacher performance and pupil learning. 4. If accountability programs are to be introduced into school districts, the programs should be designed for the improvement of instruction and not primarily to save money or dismiss teachers. 5. School districts must provide better information regarding the purposes and practices of the districts' testing programs, especially to younger, less experienced teachers. 6. Professional organizations like the National Council Teachers of English and the New York State English Council need to continue to keep teachers informed about the use and misuse of standardized testing and about the accountability movement in education.

Order No. 77-6143, 224 pages.

THE USES OF GRAMMATICAL AND RHETORICAL NORMS, PEDAGOGICAL STRATEGIES, AND STATISTICAL METHODS IN DESIGNING AND VALIDATING A COMPOSITION PLACEMENT INSTRUMENT

JOHNSON, Nancy Wehrle, D.A.
The Catholic University of America, 1977

The use of standardized tests is now being challenged. The most serious question raised is whether such tests actually measure students' knowledge relative to a particular subject matter or reflect nothing more than an ability to do well on a particular kind of test. Granting that these formulations are extremes, the issue they focus is real. Are standardized tests valid? The dissertation takes up the problem from the point of view of English composition. Specifically, it is concerned with the design of a placement test for a two-year college, but, if the design itself is sound, it may, with appropriate modification, serve as a model for other institutions.

Chapter I examines the history of placement testing in English composition from 1900 to the present. Two conclusions emerge from the survey. First, where rigorous controls are established in their administration and grading, essay tests – that is, actual student writing samples – may provide materials for a standardized test. Second, the use of these materials in the construction of a standardized test needs to be guided by norms which are public and verifiable.

Chapters II, III, and IV discuss pedagogical and subject matter considerations. The primary pedagogical consideration is whether the test can provide an accurate index of students' ability to compose grammatically correct and rhetorically effective sentences. Chapter III outlines procedures for determining what constitutes grammatical correctness, addressing itself specifically to the relation between the grader's or test designer's conception of grammar and his writing of questions. Chapter IV treats the relation between the conception of rhetoric one holds and the writing of questions. Both Chapters III and IV give examples of questions based upon students' writing samples. The contention of these two chapters, and of the chapter on pedagogy, is that clarification of ideas leads to greater precision in practice, precision that is reflected in the identification of strengths and weaknesses in actual student compositions and in the use of those compositions as materials for the construction of a placement instrument.

Chapter V is a handbook composed primarily of statistical analyses. Among the topics discussed are field testing, validity, and reliability. The handbook is also accompanied by directions for administration and scoring. Chapter VI presents follow-up studies which indicate that the test has proved a valid and valuable instrument. Chapter VII offers conclusions which relate the present test to a booklet recently published by the National Council of Teachers of English. The booklet, Common Sense and Testing in English, offers guidelines for designing standardized tests; the test described in the dissertation anticipates almost completely the recommendations of the committee representing NCTE.

Order No. 77-11,039, 172 pages.

THE DEVELOPMENT OF AN INSTRUMENT FOR ASSESSING FUNCTIONAL COMMUNICATION COMPETENCE OF SEVENTEEN YEAR OLDS

MEAD, Nancy Alexander, Ph.D.
University of Denver, 1977

The purpose of the present study was to develop an instrument for assessing the functional communication competence of seventeen year olds. The instrument was designed to be used in the National Assessment of Educational Progress, a national survey which assesses the performance of elementary and secondary students with respect to specific educational objectives.

The characteristics of the National Assessment influenced the conceptual and methodological approaches used in the present research. The study adopted a functional approach to the concept of communication ability. This conceptual framework focused on the accomplishment of communication purposes in social situations. The study used an objective-referenced approach to measurement. This method was implemented in six steps: objectives development, item development, field testing, item analysis, item selection, and tests of reliability and validity.

The study generated three products: (1) selected items, (2) tests of reliability, and (3) tests of validity. The first product was a group of items which assessed four functional categories (informing, controlling, sharing feelings and ritualizing) and related communication attitudes. The four functional categories were further divided into speaking (and nonverbal expressing) and listening (and nonverbal recognizing) perspectives. Three packages of items (each including approximately 35 minutes of testing) were selected. The selected items varied from category to category in terms of numbers, difficulty level, discrimination power, and racial/ethnic bias. The selected items are described but not included in the dissertation. The National Assessment requires that their materials be kept secure.

The second product of the study was tests of reliability of the selected items. Each category of selected items was tested for internal consistency. All of the categories either met or approached an acceptable level of reliability (criterion = .70) except the two sharing feelings categories.

The third product of the study was tests of the content validity of the selected items and construct validity of the categories. The content validity of the selected items varied from category to category. All of the selected items appeared to be an accurate measure of the objectives (face validity). However, some groups of selected items appeared to sample the domain of their respective categories better than others.

The construct validity of the category system was explored by a number of statistical comparisons. Correlations were computed between subtests scores which reflected the following relationships: (1) between speaking and listening, (2) between informing and controlling, and (3) between knowledge and related attitudes. No strong relationships were identified. This indicated that the assessment categories were distinct. However, the variety of measurement strategies used in the study (direct, indirect and affective) may have contributed to the failure to find significant correlations.

The results were discussed in relationship to the nature of the content area, functional communication competence, and the nature of the survey effort, a large-scale assessment of specific objectives. It was difficult to assess the functional areas of sharing feelings and ritualizing because they are closely tied to cultural and socio-economic norms. It was difficult to establish reliability and validity because the objective-referenced approach favors the development of a limited number of items which accurately reflect specific objectives rather than a large group of items which differentiate levels of performance within a general content domain.

Order No. 77-27,426, 183 pages.

THE INFLUENCE OF READERS ON THE HOLISTIC GRADING OF ESSAYS

NEWCOMB, Joseph Scott, Ph.D.
The University of Michigan, 1977

Chairman: LeVerne S. Collet

This study examined the effects of readers and their geographic region, sex, and race on the holistic grading of essays. Four hypotheses were stated for investigation. Hypothesis One sought to determine whether readers would account for differences in mean essay grade assignment, and variance of essay grades and whether or not a satisfactory level of inter-reader reliability would be achieved. Hypothesis Two concerned the effects of the reader region, sex, and race dimensions. Hypotheses Three and Four considered the dimensions shared between reader and writer and the effect on mean grade assignment.

Readers and essay writers were identified by region, sex, and race. Readers were trained in the holistic grading method and graded the entire sample of 192 original papers.

Analysis of variance was used to test the main effects of readers, their region, sex, and race and the interactions of these reader dimensions with the same essay dimensions and the relationship between holistic essay grades and the number and type of dimensions on which readers and essay writers were the same. The Hartley F_{max} test was used to evaluate the effect of readers on the variance of essay grades. The average inter-reader correlation coefficient and 95 percent confidence limits were used to judge the adequacy of inter-reader reliability.

Readers (considered as a random variable) differed significantly in the severity with which they assigned holistic grades and the extent to which they made use of the entire range of the grading scale. The average inter-reader correlation coefficient (.72) was significantly less than the criterion (.80) adopted for this study. No significant relationships were found between holistic essay grades and reader region, sex, and race or between holistic essay grades and the number or type of dimensions shared between readers and writers. Considerable mean essay grade difference was observed between male and female and black and white seventeen-year-old writing skill during the course of data analysis. Female writers received higher grades than male writers and white performance was rated higher than that for black writers.

Analysis of the data assuming readers as a fixed effect revealed significant differences among individual reader grading practices leading to an ability to predict their tendency to be strict or lenient graders on the basis of their geographic region, sex, and race and most combinations of these three classification variables.

Regardless of the assumption made concerning reader effect, it appears from the results of this study that the holistic essay grade which a writer receives is attributable in a large measure to the reader who assigns it rather than to the writer's ability. Order No. 77-18,082, 168 pages.

A STUDY TO DETERMINE THE CONSTRUCT VALIDITY OF THE GRAMMATIC CLOSURE SUBTEST AS A MEASURE OF LANGUAGE ABILITY USING SIGNAL DETECTION THEORY

REESE, Linda Lee, Ed.D.
Oklahoma State University, 1976

Scope of Study: The purpose of this study was to determine the construct validity of the Grammatic Closure subtest of the ITPA as a measure of language ability. Signal Detection Theory was applied to the subtest in order to look at the child as a sensor and as a decision maker. One hundred and thirty-two Eastern Oklahoma Head Start children were administered the Standard version of the subtest, a Signal Detection version, and a Restatement version of the same subtest.

Findings and Conclusions: Results were analyzed using Pearson's Product-Moment Correlation Coefficients with .05 chosen as the acceptable level of confidence. The conclusions were that the study was heavily confounded by the subjects' tendency to comply on the Signal Detection version. In spite of this, there was modest support for the validity of the Grammatic Closure subtest as a measure of language performance. Further, there was support for the fact that measuring language through the use of a Restatement test rather than performance test like Grammatic Closure might show a truer picture of the young child's language functioning.

Order No. 77-5168, 65 pages.

THE DEVELOPMENT AND EVALUATION OF A TEST OF CRITICAL LISTENING FOR USE WITH COLLEGE FRESHMEN AND SOPHOMORES

RICHARDS, Roger Allan, Ph.D.
New York University, 1976

Chairman: Professor Neil M. Postman

The purpose of this investigation was to develop an objective test of critical listening ability for use with college freshmen and sophomores and to evaluate its impact on the teaching of critical listening at the college level.

The technique of content analysis of professional literature was used to develop an analysis of critical listening to serve as the basis for the specifications of the test. This approach was based upon the premise, supported by research, that critical listening is not a single, holistic skill but rather a composite of subskills. Tabulation of frequency of use of key terms in the literature on listening, critical listening, and critical thinking, coupled with the researcher's synthesis, led to the following statement of test objectives:

The Critical Listening Test will measure the abilities to (1) Identify assumptions and inferences, (2) Recognize main ideas and arguments, (3) Identify fallacies in the use of data, (4) Judge relevance of data, (5) Recognize bias in persuasive discourse, (6) Distinguish between statements of fact and statements of opinion, and (7) Recognize errors in deductive reasoning.

A technique was developed to measure listening without the risk of contamination from requiring students to read. The researcher adopted a modified multiple-choice item in which a stem was followed by two options, either, both, or neither of which might be correct. It was found that students were able to follow a single reading of such items and to respond without confusion when neither the questions nor the responses were supplied in printed form.

A preliminary form of the test was administered for the purposes of time analysis and try-out of administration procedures. A split-half reliability of .76 was achieved. Items found to be nondiscriminating were revised for the final experimental form, which was administered by tape. The tape consists of eight passages, each read by a different person, with the investigator reading directions and test questions.

The final experimental form of the Critical Listening Test was administered to 16 classes at an urban New England community college. It was possible to include 186 students in the evaluation of the test. Claims of validity were based on a correlation of .47 with the Watson-Glaser Critical Thinking Appraisal (n=40), .64 with Verbal component of the Scholastic Aptitude Test (n-57), .47 and .41 with the McGraw-Hill Reading Test (n=90, n=57), and .49 with Brown-Carlsen Test of Listening Comprehension (n=17).

Correlations of .35 to .63 between subskill scores and total score, combined with correlations of .02 to .23 between pairs of subskills, were interpreted as supporting the hypothesis that critical listening consists of at least the seven subskills tapped by the Critical Listening Test.

The drop in reliability from .76 to .58 between the preliminary and final experimental forms raised the possibility that use of different voices may have been distracting to students and thus interfered with the measurement of critical listening. Low internal-consistency reliability was not regarded as a source of weakness of the Critical Listening Test, since it is a heterogeneous, rather than a homogeneous, instrument.

Since the results obtained agreed with other findings that college students are not highly skilled in listening, the investigation supported the claim that critical listening should be taught. It was suggested that the availability of a means of measuring the skills involved should lead to an increase in the teaching of critical listening at a time when educators are primarily concerned with accountability and with measurable behavioral objectives. Order No. 77-16,443, 211 pages.

TRAINING PRE-SERVICE TEACHERS IN USE OF STANDARDIZED LANGUAGE ARTS ACHIEVEMENT TEST: AN EXPERIMENTAL PROJECT

SHANK, Kathlene Stapfer, Ph.D.
University of Illinois at Urbana-Champaign, 1976

This study is a statistical study of the training of pre-service elementary language arts teachers in use of standardized language arts achievement tests. The pre-service elementary teachers composing the study population were given meaningful and relevant measurement experiences. Each administered a standardized language arts achievement test, scored the test, and then were given the opportunity to plan, organize, execute and evaluate a language arts program in an elementary classroom for a minimum of nine calendar weeks. Two modes of intervention were employed: training and process consultation combined with training. Training was given to two-thirds of the population. This training was directed at increasing measurement competencies. Demonstration of the competencies was encouraged and observed with additional guidance provided to one-half of the two-thirds receiving training.

The statistical data collected was analyzed to ascertain the effectiveness of the modes of intervention on the pre-service teachers' performance of evaluative role behaviors. The affects of the modes of intervention on pre-service teachers' attitudes was also statistically scrutinized. Process consultation and training were found to have affected and effected: objectively measurable knowledge of standardized achievement test, attitudes, and activities.

Increased objectively measurable knowledge of standardized achievement tests was correlated with the training and process consultation given to pre-service elementary language arts teachers. The group given neither training or process consultation did not demonstrate increased objectively measurable knowledge of standardized achievement test.

Training only and process consultation and training did affect attitudes. Affected attitudes were those toward the value of tests, kinds of information, and educational viewpoints toward worthwhile language arts content and value of evaluation.

After being given only training, pre-service elementary language arts teachers held a quizzical attitude toward the value of tests. Training and process consultation combined resulted in pre-service teachers adhering more nearly to an over-rejecting attitude toward the value of tests. The group given neither training or process consultation tended toward extremism in the form of an over-rejecting attitude toward the value of tests.

Process consultation and training and training only did result in pre-service elementary teachers believing more information was so important that they would obtain it. The pre-service elementary teachers given training and process consultation desired the greatest numbers of kinds of information. The group given neither training or process consultation felt significantly fewer kinds of information were important enough to obtain.

Process consultation and training was affective in causing pre-service elementary language arts teachers to strongly agree that evaluation can help a teacher see how effective she is in the area of language arts. Process consultation and training also resulted in pre-service elementary language arts teachers strongly agreeing that to be worthwhile, language arts content must be practically oriented.

Training and process consultation were successful in stimulating initiatory activities, discussion of test results with students, gathering of information about students, and cosmetically oriented activities. Activity stimulated activity, with the most activity, process consultation and training, generally stimulating the most activity.

Pre-service elementary language arts teachers given process consultation and training and those given only training when compared to previous pre-service elementary teachers spent more time gathering information, utilized test results more, discussed test results more with cooperating teachers and other educators, utilized a greater variety of materials, employed a greater diversity of approaches, and used standardized test results for more purposes. Pre-service elementary teachers given neither training or process consultation when compared with previous pre-service elementary teachers spent equal time and did each of the above only equally as well.

Order No. 77-9183, 193 pages.

EFFECTS OF THE CULTURAL CONTEXT OF LANGUAGE ON THE COGNITIVE PERFORMANCE OF BLACK STUDENTS

SIMPKINS, Charlesetta, Ed.D.
University of Massachusetts, 1977

Director: Professor Ellis Olim

The purpose of this study was to ascertain the effects of the cultural context of language on the cognitive performance of Black inner-city students. More specifically, this study was conducted to test the general hypothesis that when Black non-mainstream students (also referred to as inner-city students) are tested for their comprehension of reading materials in Standard American English, they will score significantly higher when the instructions for the lesson are presented in Black Vernacular than when the instructions are presented in Standard American English.

From this general hypothesis the following specific, empirical hypotheses were generated: 1. When Black non-mainstream students are instructed in Black Vernacular on how to deduce the meaning of unfamiliar words and phrases from clues in reading material, they will score significantly higher, when tested in Standard American English on their comprehension of the material, than when the instructions are given in Standard American English. 2. When Black non-mainstream students are instructed in Black Vernacular on how to deduce the meaning of figurative language from clues in reading material, they will score significantly higher, when tested in Standard American English, on their comprehension of the material, than when the instructions were given in Standard American English.

The Simpkins Test of Cultural Context (STCC) was used as the assessment instrument. The STCC was developed to assess the effect of the cultural context of language on the performance of two cognitive tasks. The two cognitive tasks, Meaning from Context and Figurative Language are incorporated into the STCC as subtests.

The review of the literature was concerned with examining two conceptual models, the difference and the deficit models, which offer explanations for the poor test performance of Black students in the public schools. In addition, the assumptions of intervention programs which have emerged from the two models were examined.

A total of 234 Black students in the seventh and eighth grades participated in the study. The students were recruited from

three schools in the Detroit Public School System which serviced Black inner-city communities.

Two one-way analyses of variance were performed on the data for the total sample and for each of the three schools participating in the study. In each case, one analysis was performed on the Meaning from Context subtest and one on the Figurative Language subtest with the instructional language (Black Vernacular vs. Standard American English) serving as the independent variable.

The overall results for the total sample were significant. On the Meaning from Context subtest, the total sample obtained a mean score of 10.195 when the explanations were given in Standard English as compared to a mean score of 11.613 on the subtest with explanations in Black Vernacular ($p < .001$, one-tailed). On the Figurative Language subtest, the total sample obtained a mean score of 5.585 when the explanations were given in Standard English as compared to a mean score of 6.544 on the subtest with explanations in Black Vernacular ($p < .012$, one-tailed).

Two major conclusions may be drawn from this study: 1. Black non-mainstream students appear to be penalized by the failure of most instructional materials to incorporate the most familiar language in the explanations and instructions of cognitive tasks which they are expected to perform. 2. When the language used in tests and other instructional materials are modified to be more compatible with the cultural-linguistic environment with which Black non-mainstream children are familiar, they will score significantly higher on these items.

Suggestions are made for further research, utilizing additional cognitive tasks, to ascertain to what extent the results of this study can be generalized to the learning of cognitive tasks in general. Order No. 77-21,510, 111 pages.

THE RELIABILITY OF RATINGS ON THE ESSAY PORTION OF THE LANGUAGE SKILLS EXAMINATION

SINGLETON, Doris Jacobs, Ed.D.
University of Georgia, 1976

Major Professors: Robert Rentz and Louis Bashaw

The objective of this study was to investigate the reliability of ratings assigned on the essay portion of the Language Skills Examination. This examination is administered to test the reading and writing skills of students attending institutions in the University System of Georgia.

Four statistical procedures were applied to four different samples of essays to investigate the reliability of the essay ratings. Reliability estimates were presented in the context of the following statistical analyses: percentage estimates of rater agreement, product-moment correlations between two sets of ratings, intraclass correlational procedures suggested by Ebel, and an intraclass correlational procedure involving calculations taken directly from a contingency table of all possible score combinations and their relative frequencies. Subsequent interpretations of the reliability coefficients were based upon the specific context, point of view, and data used for computing each reliability coefficient.

The results of this study indicate that the essays written on the Language Skills Examination are being reliably scored for the intended purposes of the test. Error was reduced in the rating process by having trained raters who concentrated on global ratings and by including multiple ratings in the final score. Order No. 77-12,416, 152 pages.

EVALUATION OF FLORIDA SECONDARY ENGLISH TEACHERS

SUTTON, Gary Alan, Ed.D.
The University of Florida, 1977

Chairman: Vincent McGuire

This study was undertaken to examine the current state of faculty evaluation procedures as they applied to English teachers in Florida's public secondary schools. As yearly evaluations are a part of Florida law, the study focused on whether these were being carried out according to the stated purpose and the letter of the law. The evaluation instruments from fifty-five of the sixty-seven state school districts were analyzed and compared. Results showed a great variety of evaluation systems, from one-page forms for evaluators to list strengths and weaknesses, to sophisticated batteries of forms to be used by administrators, teachers, students, and parents in a joint effort to assess faculty performance. Only a single district had a form designed specifically for English teachers. All others used one instrument to evaluate teachers in all areas.

Surveys were gathered from 100 secondary school English teachers from a broad representative sample of public schools. The instructors reported on the current state of faculty evaluation and made suggestions for improvement. Only 4% reported clear violations of the law due to an absence of any evaluation during the 1975-1976 school year. However, 15% reported violations due to their lack of awareness of the criteria which were to be used in their evaluations. The law requires such criteria be made known to the teacher in advance.

While the letter of the law was followed in most cases, no sincere effort at a true self-improving type of evaluation was found in the majority of schools contacted. Teachers reported evaluators with expertise in the subject area in only 32% of the cases. One or two very brief observations was the rule in the preponderance of cases studied. Only 36% of the sample were satisfied with current evaluation procedures. A total of 72% rated the theory of faculty evaluation from beneficial to extremely beneficial. When asked to rate the actual practice of faculty evaluation as they experienced it, the percentage rating evaluations as beneficial to extremely beneficial dropped to forty-one.

Teachers stated that more observations were needed for proper evaluations. They felt longer observations were needed, asked to be observed by someone with expertise in the area of English, and felt that the process should be taken more seriously. Teachers expressed concern that tenured faculty members were not as accountable as those who might lose jobs due to poor ratings. Several selected factors such as school size, district population, and the like were analyzed to see if any relationship could be seen among these factors and results of faculty evaluations. Lower paying districts were consistently rated lower by teachers. Major strengths and weaknesses of the instructors, as revealed by the evaluations, were noted.

Secondary English teachers in Florida indicated their belief in faculty evaluation. Their suggestions for the improvement of the process were included in this study.

Order No. 77-25,957, 149 pages.

DETERMINING THE INTERNAL CONSISTENCY OF ENGLISH COMPOSITIONS USING SELECTED CRITERIA

WYNN, James Howard, Ph.D.
The Louisiana State University and
Agricultural and Mechanical College, 1977

Supervisor: Professor Leonard L. Kilgore

The major purpose of this study was to determine how large a sample should be taken in order to reliably estimate student syntactical ability by using T-unit related criteria. Compositions were collected from twenty-nine students in the seventh grade and from thirty students in the tenth and twelfth grades and from thirty college upperclassmen. The mean number of words per T-unit, subordinate clauses per T-unit, clauses per T-unit, and T-units per sentence was computed for each grade and for all grades combined. The mean number of words per T-unit for the seventh grade was correlated with the mean number of words of the first T-units of all the seventh grade compositions. The mean number of words per T-unit for the seventh grade was next correlated with the mean of two means: the mean number of words for the first and second T-units from all seventh grade compositions. The mean of three means was used for a third correlation. The correlation procedure was repeated to encompass ever-increasing sample sizes from the grade's compositions. Correlation coefficients were recorded and plotted on a graph. The correlation procedure was applied in the same manner for the four T-unit criteria to grades seven, ten, twelve, and to the college students as well as a combination of the four grades.

The results of the study showed that the small samples correlated in the .80's and .90's with the mean of the whole at twenty T-units or twenty sentences, indicating that a reliable estimation of a group's skill regarding T-unit length, use of clauses, and T-units per sentence can be made by using a mean of the first twenty T-units or the first twenty sentences from all the compositions in the group.

Order No. 77-25,407, 65 pages.

Copies of the dissertations may be obtained by addressing your request to:

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106

or by telephoning (toll-free) 1-800-521-3042